UTSA
The University of Texas at San Antonio™

R Bootcamp for Urban & Social Scientists

Part 1: Intro R
January 13th, 2022

Esteban Lopez Ph.D.

Agenda

- Why data analysis?
- Lab
  - R Basics
  - Using Census and ACS data
  - My first map and Table in R

Why R?

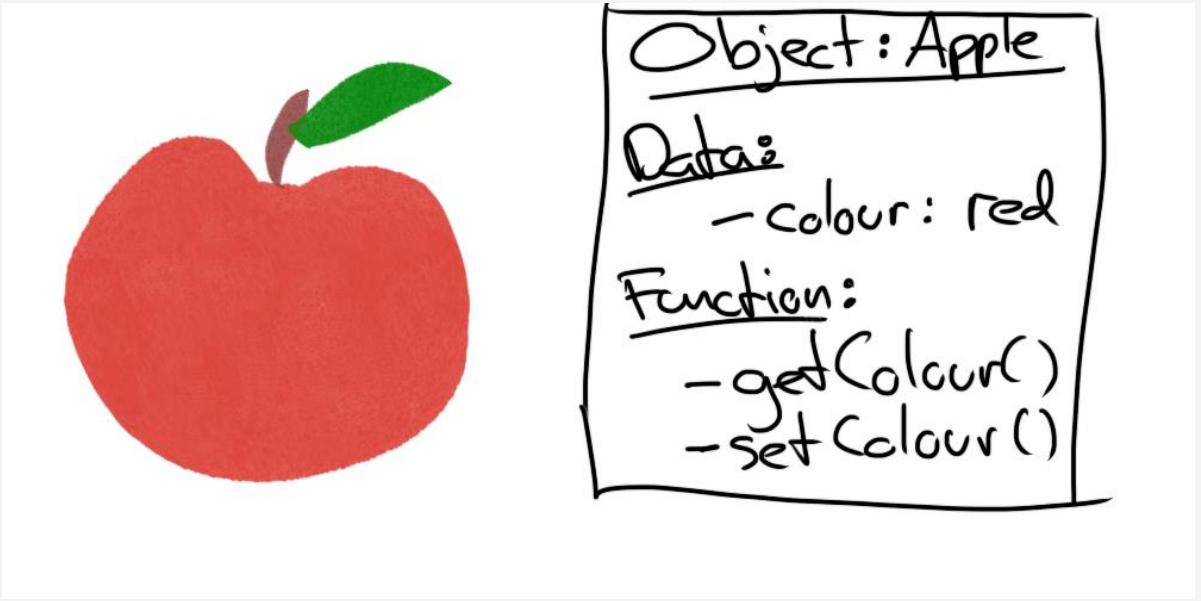
- Open source
- Great community support
- Scalable language
  - You can do *everything* in R
- A “hub” software
- At the frontier of knowledge
- Reproductible research made easy
- The ‘industry standard’ in many industries and social science research

R and R Studio

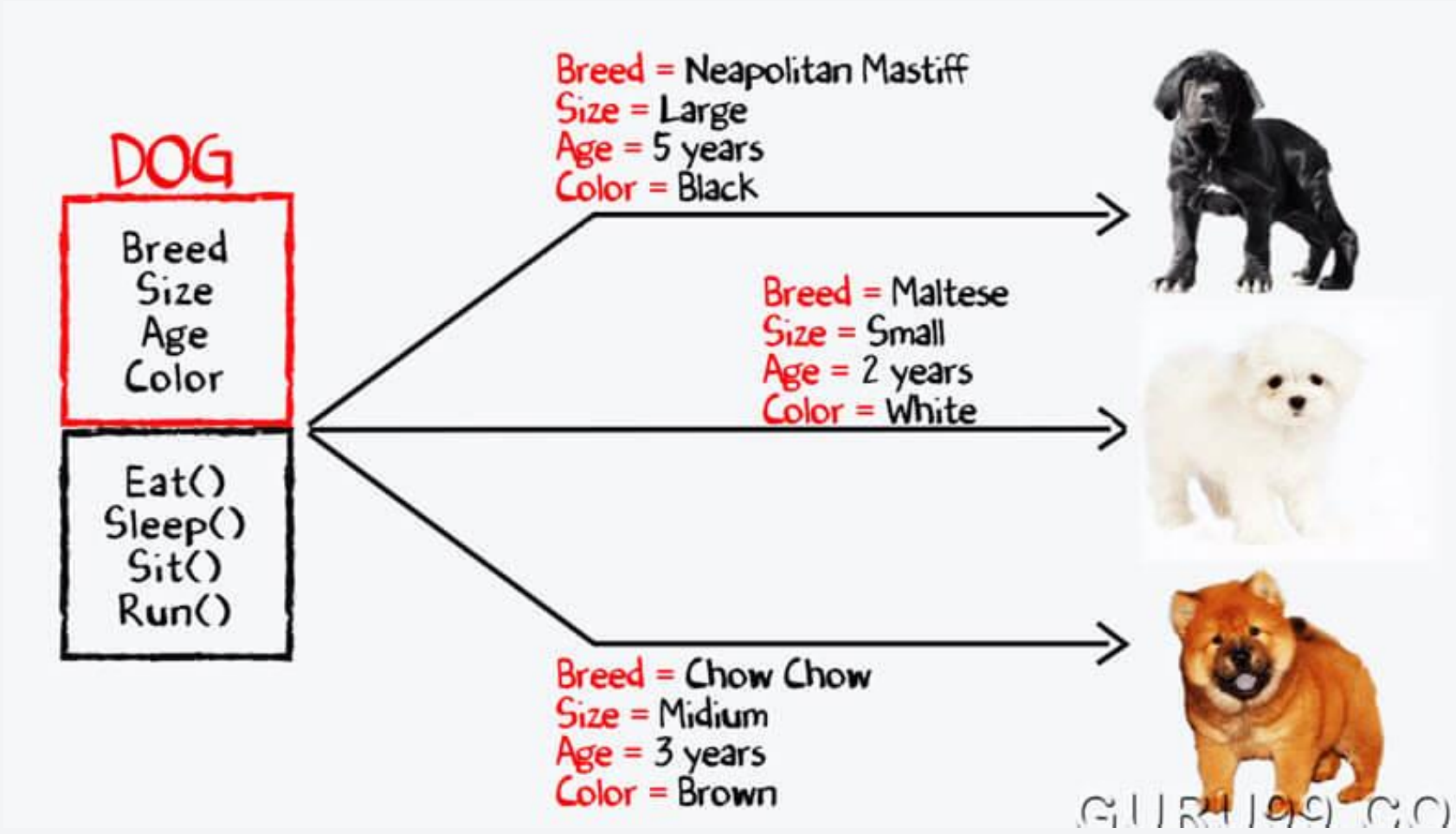
- R is the engine
- RStudio is the shell that holds and brings its capabilities to the max
  - You can do everything in R GUI,
  - But you can do it faster and in a more organized way in Rstudio IDE
- RStudio has free basic versions
  - Desktop, Cloud, Server
  - RStudio support tidyverse lenguage

Object Oriented programming

- Different from ‘task oriented’ programming
- Focused on;
  - Gather and classify data in manageable objectS (more than 1)
  - Work/manipulate objects to produce information/knowledge
- Example
  - Dog object



Object Oriented programming



R Lenguage – Sintaxis 1: Comands

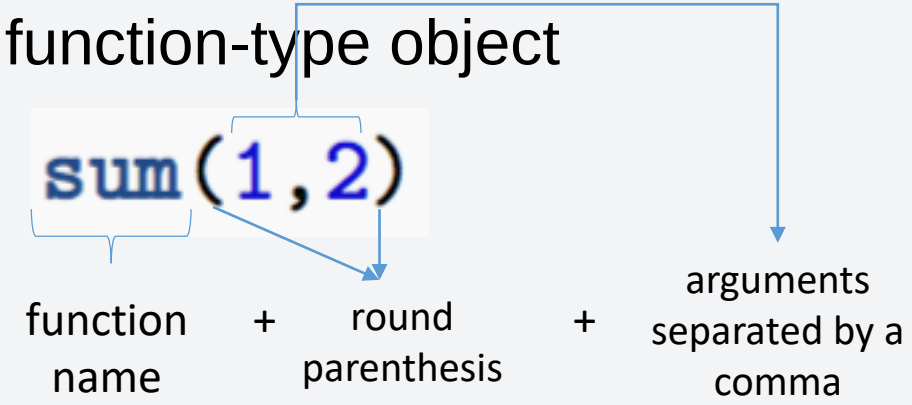
- Sign <- *asigns* a value (data structure) to an object
- *Command*: user order that R excecutes
  - You can write a command directly in the R *Console*, o write it in an R *Script* (*Source*) and the pass it to the R Console (brain/engine)
- *Console:* window where R commands are entered and output is provided (if asked for it)
- *Script: plain* text file with a set of commands to achieve a purpose
  - One command by line (recommended)
  - You can do multiple commands in each line separated by ";"
- R es *case-sensitive*
- An R command can:
  - Run something in R (without creating anything)
  - Create an object
  - Manipulate an object

R Lenguage – Sintaxis 1: Scripts

- How do you write an R Script?
  - An script it's a list of tasks executed chronologically
  - Scripts have:
    - Objects, Functions, libraries, metadata
  - Reproductibility:
    - Ability to create something and then being able to reproduce it (show how you did it)
  - Parts of a script:
    - Metadata:
    - Install & Load libraries
    - List of commands (hopefully grouped by subtasks and including comments (#))
    - Save/ export results

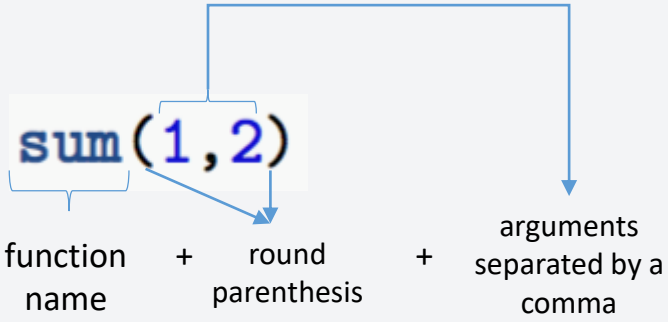
R Lenguage – Sintaxis 2: Functions

- Functions in R
  - Set of commands compacted in a function-type object
  - Parts:
    - Function name
    - Parenthesis
    - Arguments

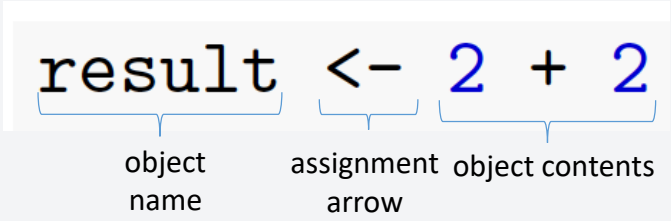


R Lenguage – Sintaxis 2: Funciones

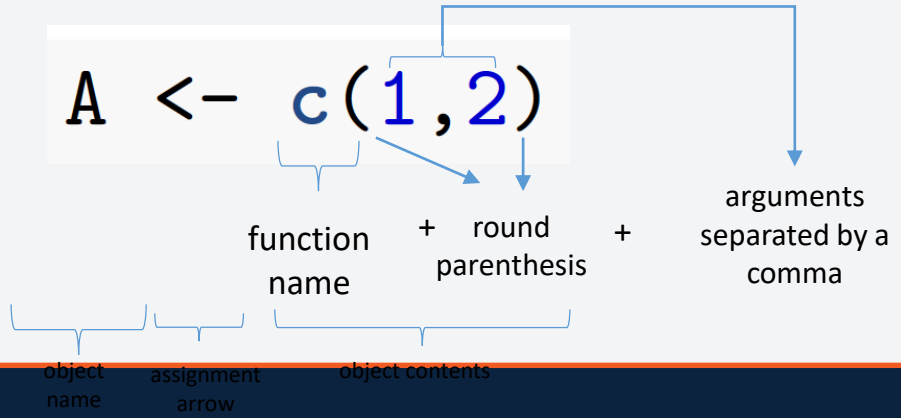
Parts of a Function



Creating an object



Creating an object using a function



R Lenguage – Sintaxis 2: Objects

- Types of Objects in R
  - There are different type of objects to store different types of data
  - Most common object types are:

Object	R command	Example
numeric	none	`a<-1`
string	`" "`	`b<- "You know nothing JS"`
list	`list()`	`l1 <- list(a,b)`
matrix	`matrix()`	`m1<- matrix(0,2,2)`
sequence	`seq()`	`sq1<-seq(1,10,1)` and `sq2<-letters[sq1]`
dataframe	`data.frame()`	`df1<-data.frame(sq1,sq2)`

  - Object Parts:
    - Elements: object contents
    - Attributes: object characteristics (what can I do with this object?)
  - Object dimensions:
    - Ways in which elements are organized within an object
      - One-dimensional – parameter, vectors
      - By-dimensional – matrices, dataframes
      - Three-dimensional – arrays
      - N-dimensionals – lists

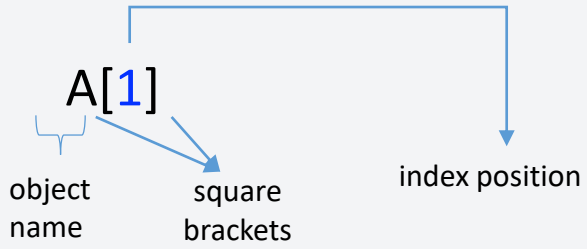
R Lenguage – Sintaxis 4: Indexing

- Indexing en R
  - How do we look for an specific element within an object?
  - Answer:
    - By position (manual)
    - By condition (filter like)
  - Sintaxis:
    - Use square brakets [ ] next to an object
    - If the object has more tan one dimensión, you need to use commas [,]

R Lenguage – Sintaxis 4: Indexing

Indexing a one-dimensional object

```
A <- c(1,2)
```



A[1] : get the first element within A

A : get ALL elements within A

A[2] : get the second element within A

A[-1] : get All elements within A, but the first

A[1:2] : get elements within A from the first to the second position

R Lenguage – Sintaxis 4: Indexing

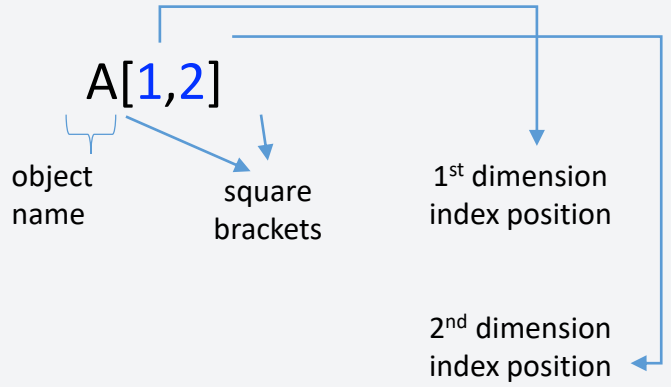
Indexing a multi-dimensional object

```
A<-matrix(1:4,2,2)
```

columns

rows

```
     [,1] [,2]
[1,]    1    3
[2,]    2    4
```



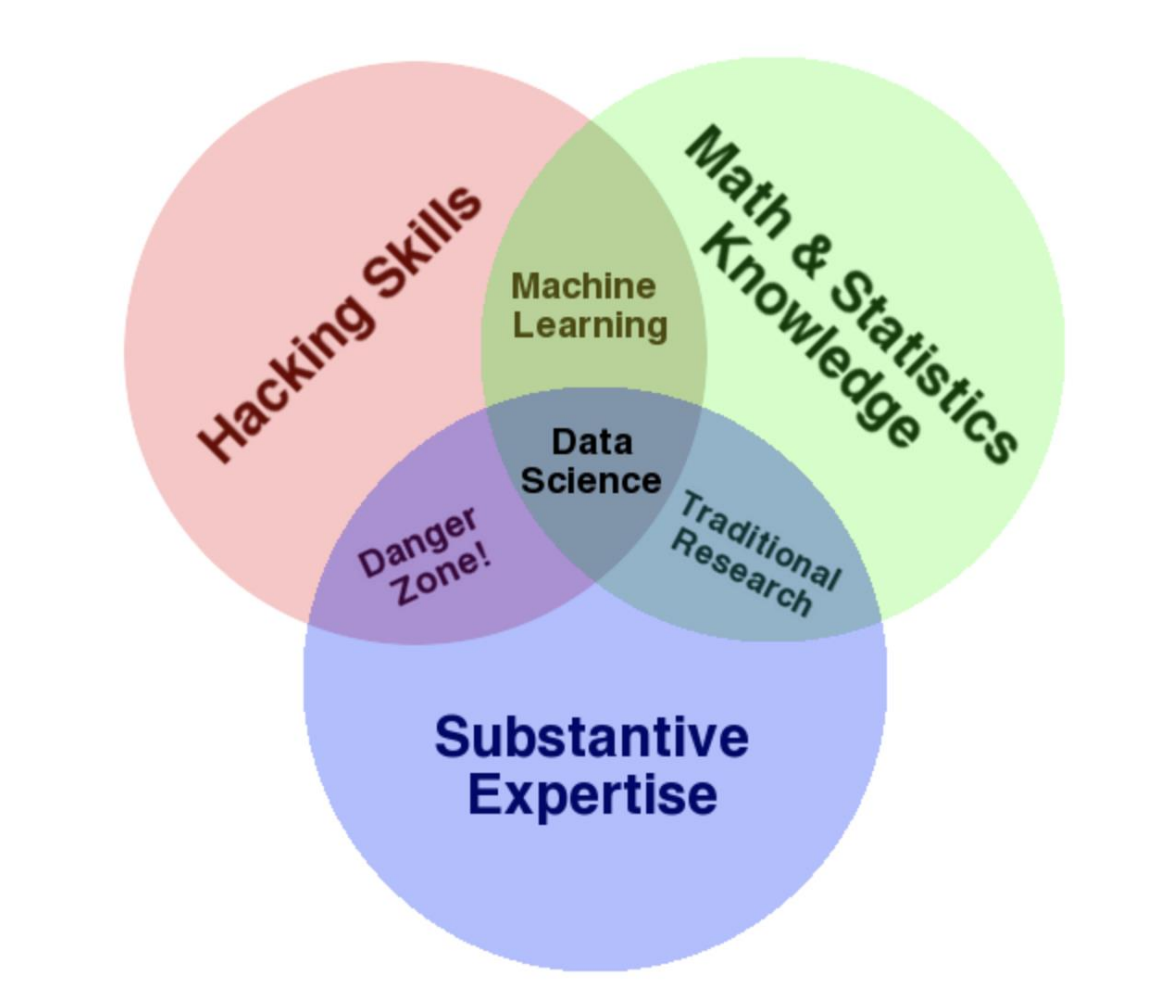
`A[1,2]` : get me the element within the object A that is located in the first row and second column

Break time

Why data analysis?

- Big data revolution
  - Data → information → Knowledge → Power
    - We need to democratize access to data
- Data driven vs. data informed Science and Policymaking
- Artificial intelligence is here
  - Be part of the change or be consumed by it?
- Smart, digital twin cities will become the norm?
  - To what extent?

Data Science



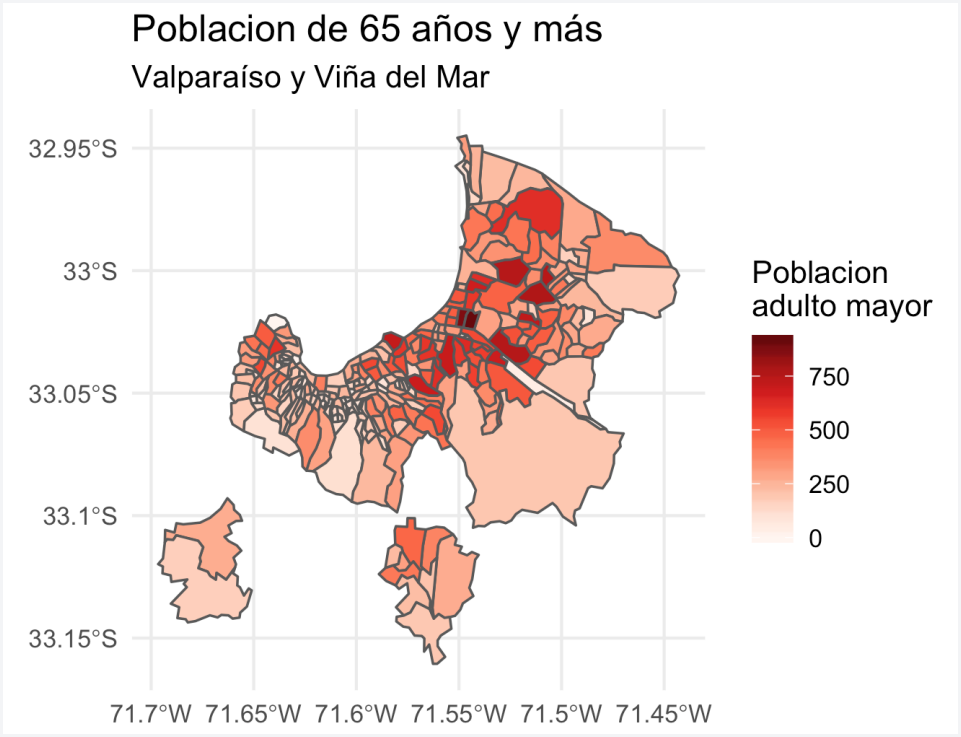
http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

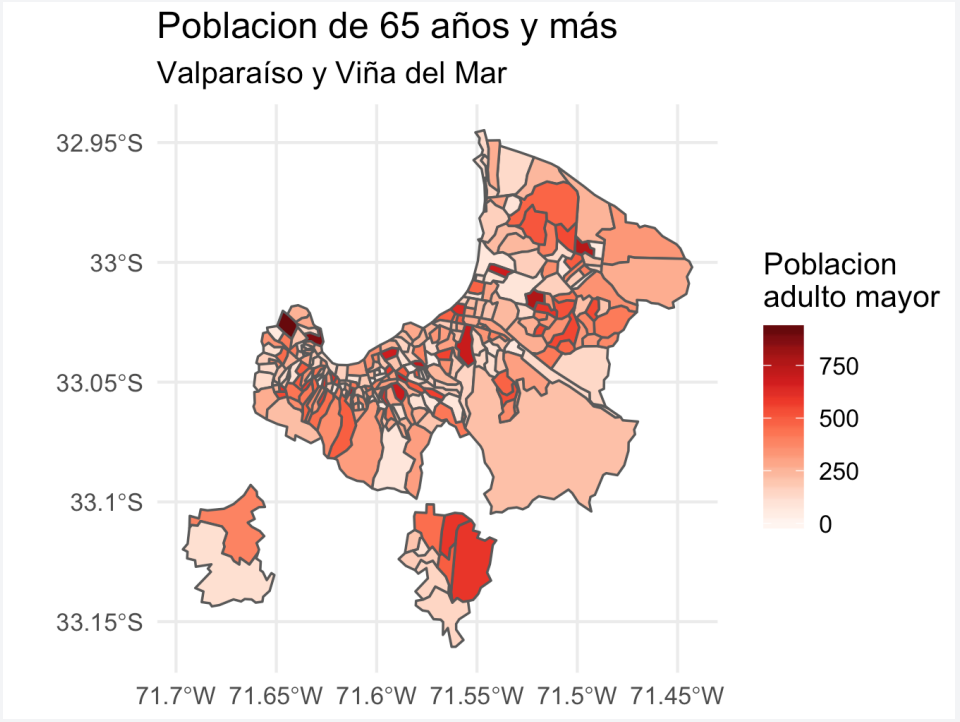
Why spatial data analysis?

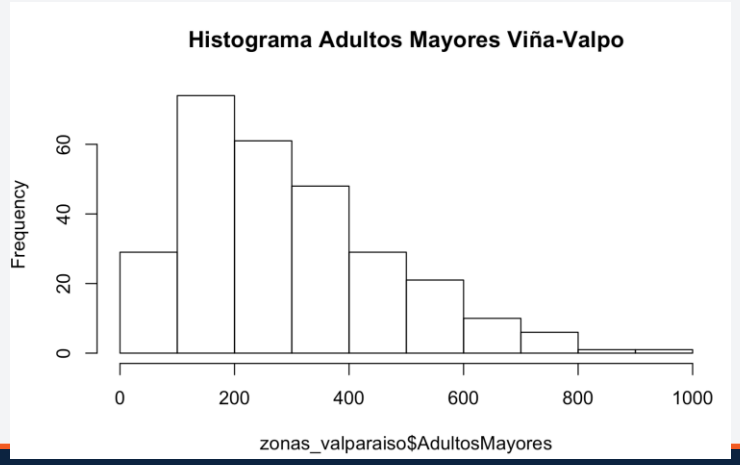
- Substantive Motivations
  - Individualism → social networks
  - Spatial externalities
  - Spatial context
- Practical motivations
  - Data: geolocation
  - Data vs. social processes mismatch
  - Spatial interpolation
  - Change of support
  - More than just mapping

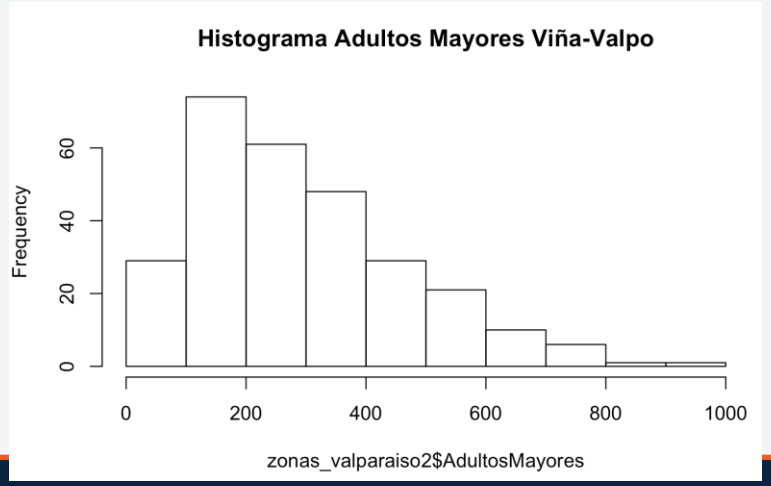
Questions of Spatial Analytics

- Where do things happen
  - Patterns, clusters, hot spots, disparities,..
- Why do things happen
  - Location decisions
- How, things that happen, affect other things (spillovers) and how context affect what happens (interaction)
- Where should things be happening/be located
  - Optimization


Poblacion de 65 años y más
Valparaíso y Viña del Mar
32.95°S
33°S
33.05°S
33.1°S
33.15°S
71.7°W 71.65°W 71.6°W 71.55°W 71.5°W 71.45°W
Poblacion adulto mayor
750
500
250
0


Poblacion de 65 años y más
Valparaíso y Viña del Mar
32.95°S
33°S
33.05°S
33.1°S
33.15°S
71.7°W 71.65°W 71.6°W 71.55°W 71.5°W 71.45°W
Poblacion adulto mayor
750
500
250
0


Histograma Adultos Mayores Viña-Valpo
Frequency
0
20
40
60
0
200
400
600
800
1000
zonas_valparaiso$AdultosMayores

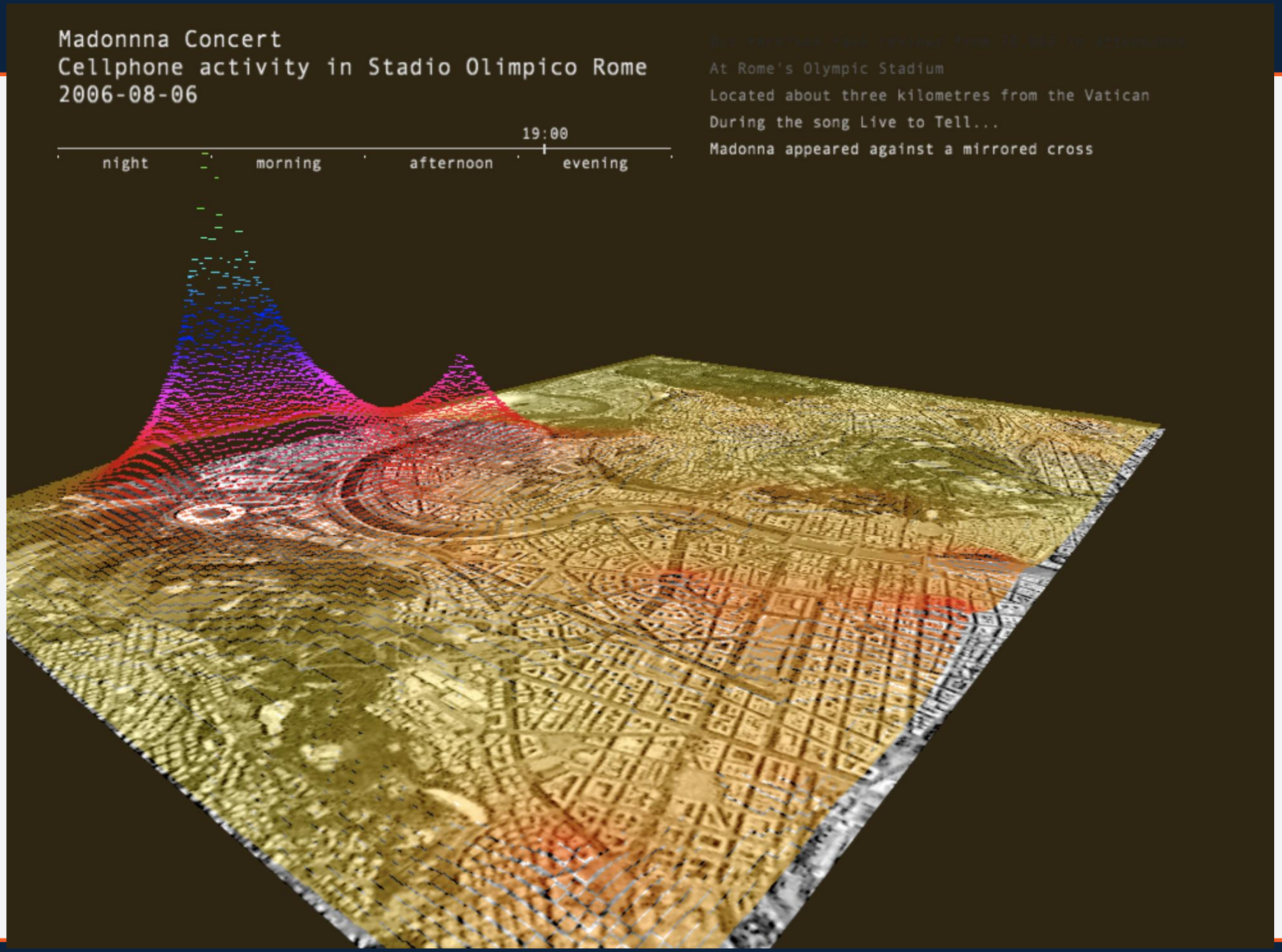

Histograma Adultos Mayores Viña-Valpo
Frequency
0
20
40
60
0
200
400
600
800
1000
zonas_valparaiso2$AdultosMayores

Big Data Issues

- Sample sizes
  - Sample vs. Population
  - Does sample size compensate for imprecision?
- Correlation is not causation

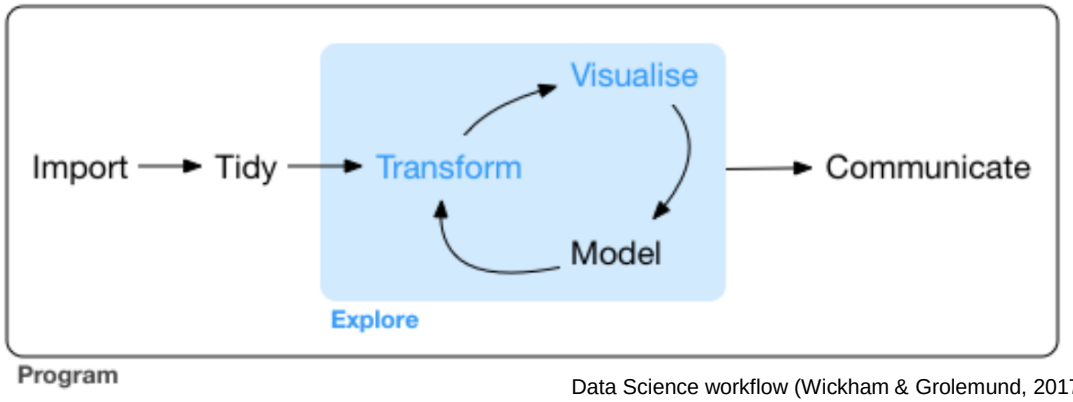
- Prediction rather than explanation

- Much newer sources of data with geolocation
  - Client data bases
  - City Sensors
  - Cell phone data



http://senseable.mit.edu/realtimerome/

Spatial Data Science



Data Science workflow (Wickham & Grolemund, 2017)

- Explicit treatment of spatial aspects
- Integration of geo-computation, spatial statistics, spatial econometrics, ESDA, spatial optimization, etc.
- 80% is data preparation (Dasu & Jhonson 2003)
  - Algorithms, data structure, workflow

utsa.edu